

# XML: de taal van de toekomst

Laurens Bloem en Marcel van Langen

Maerlant Internet Diensten - Postbus 182 - 4000 AD Tiel - Telefoon 0344-654060 - E-mail maerlant@maerlant.com

**XML. Zeker weten dat u ervan gehoord heeft. Maar wat is het precies? Waar komt het vandaan? En wat is ermee mogelijk? Dit artikel gaat in op de werking en mogelijkheden van XML. Iedereen die bekend is met internet is ook bekend met HTML (Hypertext Markup Language). HTML is de standaard opmaaktaal voor webpagina's. Met de sterke groei van internet en de mogelijkheden ervan, liep men steeds vaker aan tegen de beperkingen van HTML. Daarom ontstond steeds sterker de roep om een taal die tegemoet komt aan deze beperkingen, zonder de kracht van platformafhankelijke documenten te verliezen. Naar verwachting zal XML deze rol gaan vervullen. Reden genoeg om eens nader kennis te maken met deze taal van de toekomst.**

Trefwoorden: XML, HTML, markuptaal, DTD

## HTML: een vaste set tags

Om goed uit te kunnen leggen wat XML is en doet, is het eerst noodzakelijk kort stil te staan bij HTML. HTML is, net als XML overigens, ontwikkeld uit de SGML-standaard vanuit de behoefte om eenvoudig bestanden binnen een netwerk te kunnen koppelen. HTML staat voor HyperText Markup Language en bestaat uit een vaste set opmaakcodes (zogenaamde *tags*), met behulp waarvan structuur in een document kan worden aangebracht. Op welke manier dat moet gebeuren wordt bepaald door de instantie die de standaarden met betrekking tot het internet bewaakt en reguleert: het World Wide Web Consortium, meestal afgekort als W3C.

Bij browsers is ingebouwd hoe deze tags de informatie en het uiterlijk daarvan beïnvloeden. HTML vertelt de browser als het ware hoe de inhoud van het document moet worden weergegeven en naar welke bestemming de eventueel aanwezige verwijzingen moeten leiden. Overigens is het zo dat verschillende browserproducenten vooruit lopen op standaardisatie van tags, om zo een voorsprong te hebben op de concurrent. Bovendien ondersteunen ze eigen extra tags die niet tot de standaard behoren. In de praktijk leidt dit ertoe dat verschillende browsers een HTML-document verschillend behandelen, wat een deel van de applicatieafhankelijkheid van HTML teniet doet.

Hoewel HTML grote voordelen kent, zoals de applicatie- en platform-onafhankelijkheid, is het oorspronkelijk niet bedoeld als standaard voor de manier waarop documenten op het internet worden aangeboden. Daarvoor mist het belangrijke capaciteiten, zoals verdergaande mogelijkheden voor het

aanbrengen van een flexibele structuur en uitbreidbaarheid. Dit is vooral aan de orde nu steeds meer gebruik gemaakt wordt van koppeling van internetpagina's met databases en applicaties, en steeds minder van platte tekst.

## XML: uitbreidbare opmaaktaal

In antwoord op de groeiende behoefte aan een taal die uitbreidbaar is en geschikt om gegevens te structureren, kwam het W3C twee jaar geleden met de XML-standaard. XML staat voor 'eXtensible Markup Language', een document opmaaktaal. In vele opzichten is het nog steeds dezelfde taal als HTML, alleen is het veel diverser en biedt het veel meer mogelijkheden. Het meest wezenlijke verschil tussen beide opmaaktalen is dat bij XML de set *tags* naar believen kan worden uitgebreid. Hierdoor kan eenvoudig structuur worden aangebracht in (een reeks) documenten.

Opmaak (zoals lettertypen) wordt opgeslagen in een stylesheet via de zogenaamde 'eXtensible Stylesheet Language', kortweg XSL. Aan een browser die deze techniek ondersteunt kan verteld worden hoe een bepaalde tag binnen XML op het scherm gebracht moet worden. Daarmee is de opmaak van het document los komen te staan van de meer inhoudelijke waarden. Om een voorbeeld te geven: een bepaald segment uit een tekst bevat het antwoord op een vraag. Dit gedeelte kan met behulp van XML de inhoudelijke waarde 'antwoord' krijgen. De wijze waarop dit antwoord wordt getoond wordt beschreven in de betreffende stylesheet.

Wellicht dat een voorbeeld een en ander kan verhelderen. Dit XML-voorbeeld beschrijft een *record* dat wellicht aanwezig is in een document. Volgend op de beschrijving van het product trommel, volgen een aantal karakteristieken van het betreffende product.

```
<?XML version="1.0"?>
<!-- **** trommel **** -->
<PRODUCT>
  <product_id>98756</product_id>
  <product_naam>trommel</product_naam>
  <specification>
    <variable>kleur</variable>
    <value>blauw</value>
  </specification>
  <specification>
    <variable>grootte</variable>
    <value>klein</value>
  </specification>
</specification/>
</PRODUCT>
```

In het voorbeeld hierboven hebben we een aantal tags kunnen zien, die niet tot de standaard HTML behoort, zoals <specification> en <product>. Aangezien er bij XML geen sprake meer is van een vaste set van tags, zoals bij HTML, maar van een vrij te definiëren set is het nodig om vast te leggen welke tags gebruikt worden en wat hun functie is. Dat gebeurt in een zogenaamde DTD.

## DTD: de taal beschreven

Zonder DTD is een XML-document onmogelijk te interpreteren. De tags die gebruikt worden behoren immers niet tot een vaste set, waardoor niet eenduidig is wat elke tag betekent. Dit is overigens minder complex dan het wellicht lijkt. Als HTML niet ook een soort van DTD zou bevatten, zou het voor browsers onmogelijk zijn geweest de HTML-code juist te interpreteren. Overigens is er op dit moment een discussie gaande om ook HTML op te zetten als XML-standaard (en dus ook te voorzien van een eigen DTD).

DTD staat voor *Document Type Definition* en is een verzameling regels die samen een taal definiëren. Dit wordt gedaan via een set van zogenaamde elementen. Deze definitie beschrijft welke entiteiten er in een document bestaan, welke tags gebruikt worden en hoe deze tags een structuur kunnen vormen. De manier waarop dit alles moet worden gedefinieerd is nauwkeurig omschreven door het W3C.

De DTD kan worden opgenomen in het document zelf, of er kan in een regel verwezen worden naar de te gebruiken DTD. Aan de hand van de DTD kan het document worden gevalideerd, dat wil zeggen gecontroleerd op geldig gebruik van tags, en vervolgens geïnterpreteerd.

Voor het hierboven getoonde XML-voorbeeld zou de DTD er ongeveer als volgt uitzien:

```
<!DOCTYPE PRODUCT [  
  <!ELEMENT PRODUCT (product_id, product_name, specification*)>  
  <!ELEMENT product_id (#PCDATA)>  
  <!ELEMENT product_naam (#PCDATA)>  
  <!ELEMENT specification (variable, value)>  
  <!ELEMENT variable (#PCDATA)>  
  <!ELEMENT value (#PCDATA)>  

```

De kracht van deze methode schuilt primair in het feit dat groepen gebruikers nu een middel hebben om een aangepaste opmaaktaal te maken, die precies aansluit bij het type gegevens dat zij willen publiceren, verzenden of inlezen. Op dit moment wordt er al op vele plaatsen in de wereld gewerkt aan standaard-DTD's. Een voorbeeld hiervan is IMS (Instructional Management Systems), dat in samenwerking met Maerlant een Q&TI-DTD opstelt voor 'Question and Test Interoperability'. Met deze standaard kunnen resultaten van tests (te denken valt aan enquêtes, examens, *assessment*-tests en dergelijke) gepresenteerd worden op het internet, of zelfs uitgewisseld worden tussen programma's, zonder dat er ambiguïteit kan optreden. Dat brengt ons op een ander groot voordeel van XML: gegevensuitwisseling.

## XML: EDI van de toekomst?

XML is meer dan alleen de opvolger van HTML voor het tonen van documenten in een browser. Omdat XML het mogelijk maakt aan documenten een eigen structuur te geven, is het ook uitermate geschikt voor de uitwisseling van gegevens. En juist het eenvoudig uitwisselen van gegevens is steeds vaker een bottleneck voor de succesvolle implementatie van bijvoorbeeld keteninformatiesystemen. Zolang zender en ontvanger gebruik maken van dezelfde DTD, is een document door de ontvanger exact zo te interpreteren als de verzender het bedoeld heeft. Dit kunnen geautomatiseerde systemen zijn, met als zender bijvoorbeeld een webpagina die gegevens van een gebruiker verzamelt en verstuurt naar een ontvanger in de vorm van een applicatie. Deze applicatie kan de gegevens opslaan in een database of deze samenvoegen met andere gegevens en verwerken tot grafieken en rapporten.

De vergelijking met EDI, dat gegevens tussen mainframes uitwisselt door middel van standaarden als ANSI X12 en EDIFACT, is dan niet ver meer. En inderdaad zijn er verschillende initiatieven gaande om EDI-communicatie te baseren op de XML standaard. De voornaamste toepassing hiervan lijkt wel *e-commerce*: het bedrijven van elektronische handel. Te denken valt bijvoorbeeld aan het verzenden van transactiegegevens naar computers van banken met behulp van de XML-standaard of het rechtstreeks verzenden van bestellijsten naar computers van leveranciers. Realiteit is wel dat het nog enige tijd zal duren voor er sprake kan zijn van één standaard. Dit heeft vooral te maken met de vaak hoge investeringen die bedrijven hebben gedaan in de huidige EDI-standaarden en het groot aantal betrokkenen bij een dergelijk proces.

## XML in de praktijk

XML staat nog in de kinderschoenen en de ontwikkelingen volgen elkaar in hoog tempo op. Met name de leveranciers van software zijn momenteel bezig met een inhaalslag.

Voor gebruik van XML op grote schaal moet nog veel werk gedaan worden. De voordelen van XML bij de elektronische uitwisseling en verwerking van gegevens, zeker in relatie tot het internet, zijn echter evident. Mede omdat alle grote spelers in de relevante segmenten van de economie hun vertrouwen in de XML-standaard hebben uitgesproken, kan nu al voorzichtig gesteld worden, dat XML de taal van de (nabije) toekomst is. Meer te weten komen over XML en de mogelijkheden voor u en uw instellingen kan door contact op te nemen met Maerlant. Ook het internet is een rijke bron voor informatie. Enkele relevante sites zijn:

Algemene informatie over XML

<http://www.xml.com>

<http://www.w3c.org/XML>

Veel gestelde vragen over XML

<http://www.ucc.ie/xml/>

<http://msdn.microsoft.com/xml/general/xmlfaq.asp>

Informatie over XML in relatie tot EDI

<http://www.metronet.com/~rawlins/xmlledi.html>

<http://www.geocities.com/WallStreet/Floor/5815>

Informatie over XSL

<http://www.w3c.org/Style/XSL>